RESEARCH ARTICLE

Measurability and reliability of assessments of root length and marginal bone level in cone beam CT and intraoral radiography: a study of adolescents

1,2Ayman Al-Okshi, 1Liselotte Paulsson, 1Madeleine Rohlin, 1,3Eman Ebrahim and 1Christina Lindh

1Faculty of Odontology, Malmö University, Malmö, Sweden; 2Department of Oral Medicine and Radiology, Faculty of Dentistry, Sebha University, Sebha, Libya; 3Department of Orthodontics, Faculty of Dentistry, Sebha University, Sebha, Libya

Objectives: To evaluate measurability and reliability of measurements of root length and marginal bone level in CBCT, periapical (PA) and bitewing (BW) radiographs.

Methods: CBCT of both jaws, PA of maxillary incisors and posterior BW radiographs of 10 adolescents (mean age 13.4) were selected. The radiographs comprised part of the baseline examinations of a trial of orthodontic treatment. Six raters assessed measurability and measured root length and marginal bone level. Three raters repeated their assessments. Measurability was expressed as frequency of interpretable sites and reliability as intraclass correlation coefficient (ICC).

Results: Measurability was 100 % in CBCT and 95 % in PA of maxillary incisors for root length measurements. For marginal bone level, measurability was 100 % in CBCT, 76 % in PA and 86 % in posterior BW. Mean ICC for interrater reliability for root length measurements in CBCT was 0.88 (range 0.27–0.96 among different teeth) and 0.69 in PA of maxillary incisors. For marginal bone level measurements, mean ICC was 0.4 in CBCT, 0.38 in PA of maxillary incisors and 0.4 in posterior BW. Intrarater reliability varied among methods, root length or marginal bone level and among raters, except for root length measurements in CBCT, which presented high reliability (above 0.8) for all raters.

Conclusions: As measurability and reliability were high for root length measurements in CBCT, this may be the method of choice for scientific analyses in orthodontics. For clinical praxis, we recommend PA following the “as low as diagnostically acceptable” principle, as clinical decisions seem to be influenced only when severe root resorption occurs.


Keywords: radiography, dental; cone-beam computed tomography; humans; orthodontics; reproducibility of results

Introduction

Therapeutic interventions, such as orthodontic treatment, can be associated with adverse effects of varying severity. Most research on adverse effects of orthodontic treatment has focused on external root resorption,1–4 which is commonly expressed as a change of root or tooth length over time. In most clinical studies, this change has been assessed by means of periapical (PA) radiography of the maxillary incisors. However, this imaging method has certain shortcomings regarding image distortion and tissue overlapping, even when efforts are made to ensure accurate technique concerning receptor positioning and tube angulation. With the introduction of CBCT during the 1990s, a possibility to overcome these shortcomings opened up and over the past decades, CBCT has become a frequently used
imaging method in orthodontics. With regard to root length measurements, CBCT images underestimated root lengths of porcine teeth by less than 0.3 mm as compared to an average of 2.6 mm for PA radiographs. For repeated measurements of root length of a dry skull by one rater, errors were small in CBCT, ranging between 0.19 and 0.32 mm.

There has not been a great deal of research focused on the adverse effects on the marginal bone tissue after orthodontic treatment. According to a systematic review of orthodontics, "orthodontic treatment can cause a reduction of the bone level between teeth; the scope of this reduction, however, is so small that it lacks clinical relevance". This conclusion was based on studies using bitewing (BW) radiographs and limited to what occurs at the mesial and distal sites of the roots. Using CBCT, it was found that bone height decreases on the buccal and lingual surfaces of incisors after orthodontic treatment indicating the usefulness of CBCT imaging for analysis of marginal bone tissue.

A diagnostic test produces interpretable/uninterpretable results with varying frequencies depending on the test as well as on the diagnostic task. Frequencies of such results are reported to a limited extent although these will have an effect on test performance. This is pertinent to measurements of the tooth/root length and the marginal bone level, where failure to identify the reference points may cause uninterpretable results. One crucial reference point for these measurements is the cemento-enamel junction (CEJ), which has been reported to be uninterpretable in BW radiographs used for linear measurements of the marginal bone tissue.

In a study of root length measurements of porcine teeth made in PA radiographs and CBCT images, difference between techniques was attributed to difficulties in identifying the CEJs in the PA radiographs. Not only the diagnostic task but also characteristics of the examined sample affect the interpretation of images. In the main group of patients in orthodontic treatment, the root development and tooth eruption is ongoing. This means that the apical part of the root and marginal bone crest may be incompletely mineralized, which may cause problems when identifying reference points for measurements of root length and marginal bone level.

When assessing dental and bone tissues, the validity of the imaging method is dependent on accuracy and reliability. Investigating the agreement between and within raters provides information about the amount of error inherent in a diagnosis or score and the rater agreement may represent an “upper boundary” for diagnostic accuracy efficacy.

Knowledge concerning reliability is a prerequisite when analysing results of baseline and follow-up examinations of orthodontic treatment. To the best of our knowledge, the intra- and interrater reliability of root length and marginal bone level measurements has not been assessed for intraoral radiography and CBCT of adolescents. Furthermore, the frequency of interpretable or uninterpretable test results is an important indicator of the usefulness of an imaging method in a targeted clinical context (condition). The aim of this study was therefore to analyse the frequency of interpretable results, i.e. measurability and the reliability of measurement of root length and marginal bone level in CBCT, PA radiographs of anterior incisors, and posterior BW radiographs performed in a clinical trial of adolescents.

Methods and materials

This is a prospective study of measurability and reliability of root length and marginal bone level measurements in CBCT, PA and BW images in adolescents. The study comprises part of the baseline examinations of a clinical trial of orthodontic treatment and no additional radiographic examination was performed for the present study. The Regional Ethical Review Board, Lund, Sweden, gave ethical approval (D.no: 2014/647) to the clinical trial with the radiographic examinations, included in the present study. This study was conducted, analysed and reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies.

Sample

Adolescents with permanent teeth, crowding and tooth displacement were examined with one CBCT scan 8 x 8 cm presenting both jaws, three PA radiographs of maxillary incisors and two posterior BW radiographs in each side of the mouth. For the present study, the radiographic examinations of 10 adolescents (mean age 13.4; range 12–17) with a male-to-female ratio of 1:1, examined during March 2016 and March 2017, were selected. Table 1 presents the patient distribution, teeth and sites selected for measurement of root lengths. In the case of a premolar with two apices, the buccal root was measured. In Table 2, the patients, teeth, and sites selected for measurements of the marginal bone level are shown.

Radiographic equipment and data processing

CBCT images were obtained with 3D Accuitomo® 170 (J. Morita, Kyoto, Japan) units in two departments, using the same scanning protocol and operating at 80 kV and 3 mA. The option of 360 revolution of the X-ray source and standard acquisition mode with a field of view (FOV) of 8 cm (diameter) x 8 cm (height) and 160 µm voxel size were used to examine the upper and lower jaw together. The patients were oriented with the same plane setting controlled by the first author as a part of the study protocol.

PA and BW radiographs were obtained with a rectangular positioning device and paralleling technique in four clinics according to written instructions. The dental X-ray units, exposure parameters and imaging systems are shown in Table 3. The X-ray units were equipped with electronic timers. Prior to the radiographic
Table 1  Patient distribution and number of sites available for measurement of root lengths for CBCT and periapical radiography for each of six raters

<table>
<thead>
<tr>
<th>Patient</th>
<th>1, 3, 5, 7, 9</th>
<th>2, 4, 6, 8, 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tooth root</td>
<td>16P</td>
<td>16DB</td>
</tr>
<tr>
<td>CBCT</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Periapical radiography</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Tooth root</td>
<td>46M</td>
<td>45</td>
</tr>
<tr>
<td>CBCT</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Periapical radiography</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Patient</td>
<td>2, 4, 6, 8, 10</td>
<td>1, 3, 5, 7, 9</td>
</tr>
</tbody>
</table>

p = palatal, DB = disto-buccal, MB = mesio-buccal, D = distal, M = mesial

Table 2  Patient distribution and number of sites available for measurement of marginal bone level for CBCT, periapical, and bitewing radiography for each of six raters

<table>
<thead>
<tr>
<th>Patient</th>
<th>1, 3, 5, 7, 9</th>
<th>2, 4, 6, 8, 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tooth root</td>
<td>16P</td>
<td>16DB</td>
</tr>
<tr>
<td>CBCT</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Bitewing radiography</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Periapical radiography</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Tooth root</td>
<td>46M</td>
<td>45</td>
</tr>
<tr>
<td>CBCT</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Bitewing radiography</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Patient</td>
<td>2, 4, 6, 8, 10</td>
<td>1, 3, 5, 7, 9</td>
</tr>
</tbody>
</table>

p = palatal, DB = disto-buccal, MB = mesio-buccal, D = distal, M = mesial
examinations, one of the authors checked the radiographic equipment parameters as described by Senneby et al.4 Intraroral radiographs were assessed by one of the authors to be of acceptable quality.

10 CBCT volumes were stored in Digital Imaging and Communications in Medicine file format and prepared with i-Dixel software on a workstation. All intraoral radiographs were imported from the workstations of the four clinics to Image J software (National Institute of Health, Bethesda, MD). For all images a BARCO (MFGD 1318; BARCO, Kortrijk, Belgium) 18.10 greyscale liquid crystal display monitor with a luminance of 400 cd/m2 and resolution of 1280 × 1024 pixels was used. Linear measurement calibration of the four receptors, later used in Image J software. The reading room illumination was dimmed (below 50 lux as recommended by American Association of Physicists in Medicine Task Group 18)15 and kept constant. The reading distance was approximately 50–60 cm. There were no restrictions on reading time and zooming was allowed. All images were assessed in the same order.

First, the CBCT images were assessed, after 10 days the PA images and finally after further 10 days the BW images were assessed.

First, all raters attended a 10 min educational presentation given by the first author demonstrating the procedure using examples of root length and marginal bone level measurements on CBCT and intraoral images. During the session, the raters were given examples of a procedure similar to what they were expected to perform. The aim was to familiarise the raters with the Image J software and measurement procedure. Firstly, the raters were asked to identify the reference points. Secondly, when reference points were interpretable, they were asked to measure the root length and marginal bone level distance. Regarding marginal bone level, both mesial and distal sides of selected teeth were measured except for the maxillary canine where only the distal side was measured. All measurements recorded in millimetres were rounded to one decimal. The patient information was masked on all images.

The following definitions were applied:

- Measurability: frequency of sites with interpretable reference points for measurements of root length and marginal bone level, respectively
- Root length: distance between the mid-point between the CEJ and root apex (Figure 1). When a root was curved the measurement could be performed in two steps: first measure was the distance from the mid-point between the CEJ and the point where the root start to curve and secondly from this point to the root apex. The distances were then summed up.
- Marginal bone level: distance between CEJ and alveolar bone crest (Figure 1).

In order to calculate intrarater reliability, three raters (two dental and maxillofacial radiologists and one orthodontist) performed replicate measurements of selected sites in CBCT images (48% of available sites), in PA (73% of available sites) and in BW (51% of available sites). There were 3 weeks between the first and second sessions in order to minimise rater recall bias.

All results were collected in a computer database for analysis. For measurability, the frequency of sites where reference points were possible to identify was calculated. For analysing the reliability of each method, Intraclass correlation coefficient (ICC 2.1) with 95% confidence intervals (CI) was calculated. All statistical analyses were performed using IBM SPSS® Statistics v. 22.0 (IBM Corp., New York, NY; formerly SPSS Inc., Chicago, IL).

### Table 3 Dental X-ray units, exposure parameters, and imaging systems used in four clinics for periapical and bitewing radiography

<table>
<thead>
<tr>
<th>Clinic</th>
<th>X-ray unit (name), exposure parameters</th>
<th>Imaging system (name)</th>
<th>pixel size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental school</td>
<td>Planmeca ProX (Planmeca; Helsinki, Finland) 60kV, 7mA, 0.125 s</td>
<td>ProSensor® (Planmeca; Helsinki, Finland)</td>
<td>Pixel size 30 × 30 µm²</td>
</tr>
<tr>
<td>Hospital 1</td>
<td>Kavo, Gendex 765 DC (Kavo; Biberach/Riss, Germany) 65kV, 7mA, 0.25 s for periapical and 0.125 s for bitewing radiography</td>
<td>ProSensor® (Planmeca; Helsinki, Finland)</td>
<td>Pixel size 30 × 30 µm²</td>
</tr>
<tr>
<td>Hospital 2</td>
<td>Planmeca Intra (Planmeca; Helsinki, Finland) 60kV, 8mA, 0.160 s</td>
<td>Schick 33 (Sirona Dental, Salzburg, Austria)</td>
<td>Pixel size 15 × 15 µm²</td>
</tr>
<tr>
<td>Orthodontic clinic</td>
<td>Sirona – HELIODENT Sigma CCD (GE/DS (Sirona Dental Systems, Bernsheim, Germany) 60kV, 7mA, 0.16 s</td>
<td>Instrumentarium Imaging, Tuusula, Finland)</td>
<td>Pixel size 39 × 39 µm²</td>
</tr>
</tbody>
</table>
interpreted and measured varied among (i) imaging methods, (ii) measured object (root length or marginal bone level) and (iii) rater. For root length all available sites were interpreted and measured by all raters in CBCT and all but one site in PA of maxillary incisors. For marginal bone level, all available sites were measured in CBCT, 79% of sites (189 of 240) in PA of maxillary incisors and 86% of sites (719 of 840) in posterior BW (Table 4). Considering the six raters, measurability for root length and marginal bone level was 100% in CBCT and for root length measurements of maxillary incisors in PA except for one rater being 95%. For marginal bone level, frequencies varied among the raters for PA (range 67–95%) and for BW (range 82–88%) (Table 4).

Results

Measurability

All rater assessments were included for final analysis. As presented in Table 4, the number of sites that were

<table>
<thead>
<tr>
<th>Rater</th>
<th>Root lengths</th>
<th>Marginal bone level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CBCT</td>
<td>Periapical</td>
</tr>
<tr>
<td></td>
<td>n = 100</td>
<td>n = 20</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>100</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 4 Measurability

n = numbers of available sites for each rater, N = total number of measured sites of six ratersBW, bitewing; PA, periapical. Frequency(%) of measured sites related to number of available sites for each rater (n) for CBCT, PA and BW radiography for measurement of root length and marginal bone level. N = total number of sites measured by six raters.

Root length measurements

Interrater reliability: For all raters and roots, ICC between measurements performed in CBCT was 0.88 (CI 0.85–0.98) (Figure 2). ICC for roots of different teeth, presented in Figure 3, ranged between 0.27 and 0.96, being the highest for mandibular left second premolars and the lowest for maxillary right canines. ICCs for measurements of maxillary right central incisor and left lateral incisor in CBCT were 0.88 (CI 0.67–0.98) and 0.76 (CI 0.45–0.97), respectively (Figure 3). Corresponding ICCs for measurements of these teeth in PA were 0.64 (CI 0.28–0.94) and 0.68 (CI 0.35–0.95), respectively (Figure 3). ICC between measurements performed in PA of all four maxillary incisors was 0.69 (CI 0.52–0.83) (Figure 2). For pairwise interrater reliability, ICC of all measured roots in CBCT was above 0.85 (range 0.85–0.91). For measurements of the maxillary incisors in PA, ICC was above 0.44 (range 0.44–0.86).

Intrarater reliability: ICCs among the three raters for measurements performed in CBCT were comparable, ranging between 0.82 and 0.92 (Figure 4). For measurements of the maxillary anterior incisors performed in PA, ICCs ranged between 0.47 and 0.84 (Figure 4).

Marginal bone level measurements

Interrater reliability: For all raters and all measured sites, ICCs were 0.4 (CI 0.32–0.47) in CBCT, 0.38 (CI 0.19–0.6) in PA of the maxillary incisors and 0.4 (CI 0.25–0.55) in posterior BW images (Figure 2). For pairwise reliability, ICCs varied among the pairs of raters being between 0.34 and 0.68 for CBCT, between 0.12 and 0.66 for PA of maxillary incisors and between 0.20 and 0.66 for posterior BW images.

Intrarater reliability: ICC varied depending on imaging method among the three raters (Figure 4). For measurements in CBCT, ICC was comparable for the raters, ranging between 0.56 and 0.57. For PA of maxillary anterior incisors, ICCs ranged between 0.29 and 0.62. ICCs for posterior BW were above 0.7 for two raters, but lower for the third rater (Figure 4).
Discussion

The findings of the present study demonstrated that the frequency of sites with interpretable reference points for root length measurements in CBCT and PA of maxillary incisors was high. Measurability and reliability was lower for measurement of the marginal bone level than for that of root length, which was high. Intrarater reliability varied among methods, object measured and raters, except for the root length measurements in CBCT, which presented a high reliability for all raters. The sample was selected from images, which were obtained for a clinical trial of fixed orthodontic appliance. No additional radiographs were performed for this study resulting in a limited number of sites for intraoral images. CBCT and intraoral radiography were not compared on measurability and reliability.
A meticulous assessment of the roots and marginal bone tissue prior to orthodontic treatment provides an important baseline to evaluate changes that may occur during orthodontic treatment and may guide treatment approaches. Therefore, the diagnostic tasks chosen were measurement of root lengths and marginal bone levels of adolescents.

The spatial resolution of CBCT mainly depends on receptor physical pixel size. Other factors such as projection geometry, scatter radiation, detector motion blur, focal spot size, number of basis projections and reconstruction algorithm may affect the image resolution. Additionally, contrast resolution is important in order to distinguish between structures with minor differences in attenuation and to present them with different grey levels. The CBCT unit used offered four predefined scan modes with possibilities to change the mA- and kV-settings and dimension of FOV. As recommended by Al-Okshi et al., full rotation with 80 kV and modified mA to 3 according to patient size was used to obtain optimal subjective image quality for assessment of periodontal structures. As a low exposure (3 mA) was used, full rotation was chosen in order to improve the signal-to-noise ratio. The size of FOV influences the image as a large FOV may reduce contrast-to-noise ratio resulting in decreased visibility of anatomical structures. Although the root apex and the marginal bone level can be subtle structures in adolescents, we selected an 8 × 8 cm FOV to capture the upper and lower jaw during the same scanning in order to reduce patient radiation dose. Another influential factor when imaging subtle anatomical structures is the voxel size—the smaller the voxel size, the higher the spatial resolution but with higher image noise, which may have influenced the reliability of the measurements of the marginal bone level.

Measurability is one way of simultaneously expressing image quality and the usefulness of an imaging method. Measurability adds to visual grading analysis of the images, which is commonly used to evaluate image quality, as it demands the rater to perform a clinically relevant task. The included patients comprised adolescents, who will have fixed orthodontic appliances and undergo several radiographic examinations during their treatment. Although all raters stated that all measurements were possible to perform in CBCT, it may be expected that this will not be the case in subsequent radiographic examinations. Then, the patients will have metal/radiopaque material on their teeth, which increases the risk that a case will be scored as not interpretable in CBCT. Thus, a lower measurability is to be expected in some the follow-up examinations of orthodontic treatment performed with CBCT. All but one root length was measured in PA of the maxillary incisors whilst the frequency of measurable sites of the marginal bone level in PA was lower and varied considerably among the raters. The frequency of uninterpretable sites for measurement of the marginal bone level in PA of the maxillary incisors (about 20%) corresponds to that found in a study of adult patients. A common reason for uninterpretable sites is said to be the difficulty in identifying the CEJ. This did not seem to be the case for PA in the present study as all but one measurement of root length, which were based on the CEJ, were performed. Instead it may be attributed to difficulties in identifying the other reference point, i.e. the marginal bone level. One explanation for this may be that some PA images were slightly distorted as it is difficult to perform PA of the maxillary incisors with optimal quality in adolescents with crowded teeth. For posterior BW the lower measurability may be due to the fact that...
the marginal bone septa were incompletely mineralized. The reported frequency of unacceptable image quality for assessment of the marginal bone tissue in posterior BW varies between 6 and 29% in previous studies of adult patients. As far as we know, there is no study of measurability of intraoral radiographs of adolescents. When planning clinical trials and analyzing treatment outcomes, results on measurability are important to take into account. If results on measurability are reported to a limited extent and simply removed from the analysis, it may lead to biased analyses of the diagnostic modalities and to incorrect reports of treatment outcomes.

Accuracy is a cornerstone in assessing the efficacy of diagnostic methods. When a reference standard is not available, reliability studies can be used to indicate the upper bound of accuracy. In other words, the reliability estimations are useful in determining the extent to which the inaccuracy of a system is due to decision-making errors. Unfortunately, reliability studies are generally neglected and do not appear in the different stages of evaluating studies of diagnostic methods or in studies where diagnostic methods are used to evaluate treatment outcomes. Changes in dental or bone tissues caused by different treatment methods are often assessed by comparing measurement results from the baseline examination with those of follow-up examinations. Then, accuracy may not be the issue but rather the reliability of the method used to assess any change. In studies of orthodontic treatment, the measurement errors of baseline and follow-up examinations should fall short of the assessed change of root length and marginal bone level. Furthermore, it is important to be aware of several raters’ performance.

In the present study, interrater reliability for six raters’ measurements of root length in CBCT was high (ICC 0.88) and so was intrarater reliability (ICC above 0.8). Although direct comparison with previous studies on reliability of root length measurements in CBCT is not pertinent, our results are broadly in agreement. The fact that interrater reliability varied among roots of different teeth, should be considered when selecting roots for analysis of treatment outcomes. In the previous studies of root length measurements in CBCT, the cusp/incisal edge represented one of the reference points. As the cusp/incisal edge may change over time due to attrition, we preferred to use the CEJ as the reference point as it is more stable when comparing root lengths. Interrater reliability of root length measurement performed in PA of anterior incisors was rather high (ICC 0.69) but lower than that reported for two raters (ICC above 0.90) by Lago et al for inter and intrarater reliability.

Overall, also in CBCT, reliability for measurement of the marginal bone level was low. The results for posterior BW were somewhat unexpected as BW is recommended for assessment of the marginal bone tissue. The low reliability may be due to the sample, which consisted of young adolescents. The delineation of the marginal bone is therefore imaged more vaguely, especially around second premolars and second molars.

Conclusions

Our results on high measurability in CBCT for measurements of root lengths and marginal bone level as well as high reliability for root length measurements further add to the results of previous clinical studies indicating that CBCT may be the best choice for scientific analyses. But for clinical praxis, we recommend intraoral radiography as preventive measures such as the use of lower forces, resting periods, and shortening of treatment time are not taken until root resorption from 2 mm up to 1/3 of the root length is diagnosed. This is in accordance with the “as low as diagnostically acceptable” principle, as the radiation dose for intraoral radiography is lower than that of a CBCT examination. Thus, a radiographic method must be justified for each patient, both regarding radiation dose and diagnostic image quality, which is emphasised in the SEDENTEXCT guidelines on CBCT for Dental and Maxillofacial Radiology.

Acknowledgement

This study project was supported by research funds for Oral Health Related Research by Region Skåne, Sweden [OFR5512901].

References

Reliability of root length and marginal bone level measurement
Al-Oksli et al

9 of 9

during orthodontic treatment. Angle Orthod 2010; 80: 466–73. doi: https://doi.org/10.2319/072909-427.1